University of Groningen

Radio interferometric calibration via ordered-subsets algorithms: OS-LS and OS-SAGE calibrations

Kazemi, S.; Yatawatta, S.; Zaroubi, S.

Published in:
Monthly Notices of the Royal Astronomical Society

DOI:
10.1093/mnras/stt1229

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Kazemi, S., Yatawatta, S., & Zaroubi, S. (2013). Radio interferometric calibration via ordered-subsets algorithms: OS-LS and OS-SAGE calibrations. *Monthly Notices of the Royal Astronomical Society*, *434*(4), 3130-3141. https://doi.org/10.1093/mnras/stt1229

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **434,** 3130–3141 (2013) doi:10.1093/mnras/stt1229
Advance Access publication 2013 July 26

Radio interferometric calibration via ordered-subsets algorithms: OS–LS and OS–SAGE calibrations

S. Kazemi,[1]⋆ S. Yatawatta[2] and S. Zaroubi[1]

[1]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[2]*ASTRON, Postbus 2, NL-7990 AA Dwingeloo, the Netherlands*

Accepted 2013 July 3. Received 2013 June 29; in original form 2012 November 13

ABSTRACT

The main objective of this work is to accelerate the maximum likelihood (ML) estimation procedure in radio interferometric calibration. We introduce the ordered-subsets–least-squares (OS–LS) and the ordered-subsets–space alternating generalized expectation (OS–SAGE) radio interferometric calibration methods, as a combination of the OS method with the LS and SAGE maximization calibration techniques, respectively. The OS algorithm speeds up the ML estimation and achieves nearly the same level of accuracy of solutions as the one obtained by the non-OS methods. We apply the OS–LS and OS–SAGE calibration methods to simulated observations and show that these methods have a much higher convergence rate relative to the conventional LS and SAGE techniques. Moreover, the obtained results show that the OS–SAGE calibration technique has a superior performance compared to the OS–LS calibration method in the sense of achieving more accurate results while having significantly less computational cost.

Key words: methods: numerical – methods: statistical – techniques: interferometric.

1 INTRODUCTION

Radio interferometry is the technique of combining and correlating signals from two or more separate antennas to observe the target astronomical object with a resolution determined not by the size of a single antenna but by the area covered with all the incorporated antennas. Therefore, a much better angular resolution can be achieved using radio interferometers with multiple antennas instead of single dishes (Thompson, Moran & Swenson 2001; Burke & Graham-Smith 2009).

The main objective of designing the new generation of radio interferometers, such as the Square Kilometer Array (SKA),[1] the Murchison Widefield Array (MWA),[2] the Precision Array to Probe Epoch of Reionization (PAPER),[3] the 21-cm Array (21CMA),[4] the Hydrogen Epoch of Reionization Array (HERA),[5] the Long Wavelength Array (LWA)[6] and the Low Frequency Array (LOFAR),[7] with the ability to collect enormous amounts of data, is improving the sensitivity, resolution and frequency coverage of observations. Therefore, to deliver their scientific goals, there is a need for processing a large amount of data and for upgrading the accuracy as well as the processing time of the existing calibration techniques.

Propagation medium and the receivers' effect in radio interferometric data are initially unknown and have to be calibrated and corrected before imaging. Self-calibration (Pearson & Readhead 1984) estimates the maximum likelihood (ML) estimate of the unknowns utilizing only the measurements, and due to its high accuracy, it has become the method of choice, as in this paper, for calibrating the new generation of radio synthesis arrays.

In the presence of additive Gaussian noise, calibration is performed as a non-linear least-squares (LS) optimization that calculates the ML estimation using iterative-gradient-based methods such as Levenberg–Marquardt (LM) method (Levenberg 1944; Marquardt 1963). However, the LS calibration suffers from a very low convergence rate because the parameters must be updated simultaneously on a complete data space. Solving for a large number of unknowns, the Jacobian computation corresponding to the applied-gradient-based method is considerably costly. This makes the LS calibration impractical for calibration of giant radio telescopes like SKA with thousands of receivers.

The convergence rate and computational efficiency of calibration are significantly improved by the recently proposed space alternating generalized expectation (SAGE) maximization calibration technique (Yatawatta et al. 2009; Kazemi et al. 2011). SAGE method (Fessler & Hero 1994) is a specific version of the expectation maximization (EM) algorithm (Feder & Weinstein 1988) which

⋆ E-mail: kazemi@astro.rug.nl
[1] http://www.skatelescope.org
[2] http://www.mwatelescope.org
[3] http://astro.berkeley.edu/~dbacker/eor
[4] http://21cma.bao.ac.cn
[5] http://www.reionization.org
[6] http://lwa.unm.edu
[7] http://www.lofar.org

© 2013 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

converges even faster than the conventional EM (Fessler & Hero 1993). The SAGE algorithm partitions the data space to smaller 'hidden' data spaces and at every iteration, it alternates between updating parameters on some or all of them. Obtaining the ML estimate for the parameters of these small data spaces, which carry less information compared to the complete data space, provides SAGE algorithm with a superior accuracy as well as lower computational cost compared to the LS technique. Nevertheless, there is still a need for improving the speed of calibration process especially for radio synthesis arrays such as LOFAR and SKA.

In this paper, we utilize the ordered-subsets (OS) algorithm for accelerating the speed of calibration. The well known OS algorithm accelerates the convergence rate of iterative ML estimations and has been widely used in medical imaging (Hudson & Larkin 1994; Erdogan & Fessler 1999). This method decomposes the objective (likelihood) function to several sub-objective functions and updates the parameters by using the gradient of one, or some, of the sub-objective functions as an approximation to the original objective function's gradient. At the initial iterations, when the parameters are far from the optimum point, these approximations are quite reasonable since the gradient is only an approximation at those stages. Thus, they can be efficient substitutions for the gradient of the original cost function and considerably accelerate the computations of the OS algorithm. However, it must be taken into account that the highest accuracy that OS methods can achieve is the same as the one which could be obtained by the conventional (non-OS) techniques. Close to the optimal solution, OS methods generally do not converge but rather become stuck at a sub-optimal limit cycle of as many points as there are sub-objective functions. Therefore, if OS method becomes globally convergent (Hudson & Larkin 1994; Ahn 2004), it maintains exactly the same accuracy of the convergent non-OS methods.

This paper is organized as follows. In Sections 2 and 3, we present the general data model of radio interferometric calibration and the classical LS and SAGE calibration techniques. In Section 4, we introduce the OS–LS and OS–SAGE calibration techniques in order to cut down the processing time of the conventional LS and SAGE calibration methods. The computational advantages of applying the OS-type calibrations instead of the conventional methods are also shown. For the ML estimations, the LM method is applied. At the end of Section 4, we show an application of OS calibration to accelerate computations when calibrating for an individual data sample. The technique is based on partitioning data over baselines and hence could also be useful in speeding up the calibration procedure of radio telescopes with a large number of receivers. In Section 5, we give results based on simulations to demonstrate the superior convergence rate of the OS calibration schemes compared to the non-OS ones. Finally, we draw our conclusions in Section 6.

The following notations are used in this paper: bolditalic, lowercase letters refer to column vectors, e.g., $\boldsymbol{y}$. Uppercase san serif letters refer to matrices, e.g.,$\mathbf{C}$. The transpose, Hermitian transpose and conjugation of a matrix are presented by $(.)^{\mathrm{T}}$, $(.)^{\mathrm{H}}$ and $(.)^{*}$, respectively. The matrix Kronecker product is denoted by $\otimes$. $\mathbb{R}$ is the set of Real numbers. E{} denotes the statistical expectation operator. The real and imaginary parts of complex quantities are shown by $\mathfrak{Re}$ and $\mathfrak{Im}$, respectively.

2 CALIBRATION DATA MODEL

In this section, the general measurement equation of a polarimetric radio interferometer is presented. For some introduction to radio polarimetry and calibration the reader is referred to Hamaker, Bregman & Sault (1996) and Hamaker (2006).

Consider a radio interferometer with N antennas which observes K uncorrelated sources. The induced voltage at antenna p, $\tilde{\boldsymbol{v}}_{pl}$, due to radiation of the lth source, $\boldsymbol{e}_l$, is given by $\tilde{\boldsymbol{v}}_{pl} = \tilde{\mathbf{J}}_{pl}\boldsymbol{e}_l$, where $\tilde{\mathbf{J}}_{pl}$ is the complex 2×2 Jones matrix (Hamaker et al. 1996) corresponding to the sky and instrumental corruptions of the signal.

The total signal obtained at antenna p, $\boldsymbol{v}_p$, is a linear superposition of K such signals plus the antenna's thermal noise. After correcting for geometric delays and the instrumental effects, the pth antenna voltage is correlated with the other $N-1$ antennas voltages. The correlated voltages $\mathrm{E}\{\boldsymbol{v}_p\boldsymbol{v}_q^{\mathrm{H}}\}$, referred to as *visibility* (Hamaker et al. 1996) of baseline $p-q$ is given by

$$\mathbf{V}_{pq} = \mathbf{G}_p\left(\sum_{l=1}^{K}\mathbf{J}_{pl}\mathbf{C}_l\mathbf{J}_{ql}^{\mathrm{H}}\right)\mathbf{G}_q^{\mathrm{H}} + \mathbf{N}_{pq}, \tag{1}$$

where $\mathbf{N}_{pq}$ is the baseline's additive noise and $\mathbf{C}_l = \mathrm{E}\{\boldsymbol{e}_l\boldsymbol{e}_l^{\mathrm{H}}\}$ is the lth source *coherency* matrix (Hamaker et al. 1996; Born & Wolf 1999). The errors common to all directions (mainly the receiver delay and amplitude errors) are given by $\mathbf{G}_p$ and $\mathbf{G}_q$. We assume that an initial calibration, at a finer time and frequency resolution, is performed to estimate $\mathbf{G}_p$s (direction-independent effects). Then, the corrected data are obtained as

$$\widetilde{\mathbf{V}}_{pq} = \mathbf{G}_p^{-1}\mathbf{V}_{pq}\mathbf{G}_q^{-\mathrm{H}}, \tag{2}$$

where $\widetilde{\mathbf{V}}_{pq}$ are the visibilities after correction for effects common to all directions. The remaining errors are unique to a given direction, but residual errors in $\mathbf{G}_p$s are also absorbed into these errors, which are denoted by $\mathbf{J}_{pl}$ in the usual notation. The vectorized form of corrected visibilities is given by

$$\boldsymbol{v}_{pq} \equiv \mathrm{vec}(\widetilde{\mathbf{V}}_{pq}) = \sum_{l=1}^{K}\boldsymbol{s}_{pql} + \boldsymbol{n}_{pq}, \tag{3}$$

where $\boldsymbol{s}_{pql} = \mathbf{J}_{ql}^{*} \otimes \mathbf{J}_{pl}\mathrm{vec}(\mathbf{C}_l)$ and $\boldsymbol{n}_{pq} = \mathrm{vec}(\mathbf{G}_p^{-1}\mathbf{N}_{pq}\mathbf{G}_q^{-\mathrm{H}})$. The unknowns of the calibration problem are the real and imaginary parts of the Jones matrices complex elements:

$$\boldsymbol{\theta} = [\mathrm{vec}(\mathfrak{Re}\{\mathbf{J}_{11}\})^{\mathrm{T}}\ \mathrm{vec}(\mathfrak{Im}\{\mathbf{J}_{11}\})^{\mathrm{T}}\ \mathrm{vec}(\mathfrak{Re}\{\mathbf{J}_{12}\})^{\mathrm{T}}\cdots]^{\mathrm{T}},$$

and therefore, $\boldsymbol{\theta} \in \mathbb{R}^{8KN\times 1}$.

Consider a data set of τ time and frequency samples that form a small enough time and frequency interval over which $\boldsymbol{\theta}$ is invariant. Stacking up the real and imaginary parts of the instrument's visibilities and noise vectors in $\boldsymbol{y} = [\mathfrak{Re}\{\boldsymbol{v}_{12}^{\mathrm{T}}\}\ \mathfrak{Im}\{\boldsymbol{v}_{12}^{\mathrm{T}}\}\ \mathfrak{Re}\{\boldsymbol{v}_{13}^{\mathrm{T}}\}\ \cdots]^{\mathrm{T}}$ and $\boldsymbol{n} = [\mathfrak{Re}\{\boldsymbol{n}_{12}^{\mathrm{T}}\}\ \mathfrak{Im}\{\boldsymbol{n}_{12}^{\mathrm{T}}\}\ \mathfrak{Re}\{\boldsymbol{n}_{13}^{\mathrm{T}}\}\ \cdots]^{\mathrm{T}}$, respectively, the general measurement equation becomes

$$\boldsymbol{y} = \sum_{l=1}^{K}\boldsymbol{s}_l(\boldsymbol{\theta}) + \boldsymbol{n}. \tag{4}$$

In (4), $\boldsymbol{s}_l(\boldsymbol{\theta}) = [\mathfrak{Re}\{\boldsymbol{s}_{12l}^{\mathrm{T}}\}\ \mathfrak{Im}\{\boldsymbol{s}_{12l}^{\mathrm{T}}\}\ \mathfrak{Re}\{\boldsymbol{s}_{13l}^{\mathrm{T}}\}\ \cdots]^{\mathrm{T}}$. $\boldsymbol{y}$, $\boldsymbol{n}$ and $\boldsymbol{s}_l$ are vectors of size $4\tau N(N-1)$, and the noise vector $\boldsymbol{n}$ is assumed to be white Gaussian. Calibration is the ML estimation of the unknown parameter vector $\boldsymbol{\theta}$ from (4). Note that calibration methods could also be applied to the uncorrected visibilities of (1) to estimate $\mathbf{G}_p$ and $\mathbf{G}_q$ errors as well. Moreover, having a large enough N and small enough K, there will be enough constrains to solve for the $8KN$ unknown parameters of $\boldsymbol{\theta}$ using the $4\tau N(N-1)$ measurements of $\boldsymbol{y}$.

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

3 THE LS AND SAGE CALIBRATION METHODS

In this section, both the LS and SAGE calibration algorithms are briefly outlined. The OS scheme is applied to both methods.

3.1 LS calibration

Since the noise vector $\boldsymbol{n}$ in the calibration data model (4) is assumed to be white Gaussian, LS calibration method estimates the ML estimate of $\boldsymbol{\theta} \in \mathbb{R}^{8KN\times 1}$ by minimizing the sum of squared errors:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\boldsymbol{y} - \sum_{l=1}^{K} \boldsymbol{s}_l(\boldsymbol{\theta})||^2. \tag{5}$$

Gradient-based optimization techniques are used for solving (5). Among those, the LM method (Levenberg 1944; Marquardt 1963) is one of the most robust in the sense that it mostly converges to a global optimum. Defining the cost function $\phi(\boldsymbol{\theta}) = \boldsymbol{y} - \sum_{l=1}^{K} \boldsymbol{s}_l(\boldsymbol{\theta})$, where $\phi(\boldsymbol{\theta}) \in \mathbb{R}^{4\tau N(N-1)\times 1}$, and initializing the starting point $\widehat{\boldsymbol{\theta}}^1$, the LS calibration method via LM algorithm is outlined as follows:
for every iteration $k = 1, 2, \ldots$ until an upper limit or convergence of $\widehat{\boldsymbol{\theta}}^k$,

Calculate $\widehat{\boldsymbol{\theta}}^{k+1} \in \mathbb{R}^{8KN\times 1}$ using LM algorithm as

$$\widehat{\boldsymbol{\theta}}^{k+1} = \widehat{\boldsymbol{\theta}}^k - (\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\nabla}_{\boldsymbol{\theta}} + \lambda\mathbf{H})^{-1}\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{\mathrm{T}}\phi(\boldsymbol{\theta})|_{\widehat{\boldsymbol{\theta}}^k}. \tag{6}$$

endfor

In (6), $\boldsymbol{\nabla}_{\boldsymbol{\theta}} = \frac{\partial}{\partial\boldsymbol{\theta}}\phi(\boldsymbol{\theta})$, λ is the damping factor (Lampton 1997), and $\mathbf{H} = \mathrm{diag}(\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\nabla}_{\boldsymbol{\theta}})$ is the diagonal of the Hessian matrix. The sizes of the Jacobian $\boldsymbol{\nabla}_{\boldsymbol{\theta}}$ and the linear system solved in (6) are $4\tau N(N-1)\times 8KN$ and $8KN$, respectively. Consequently, the cost of computing $\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\nabla}_{\boldsymbol{\theta}}$ is $\mathcal{O}((8KN)^2 \times 4\tau N(N-1))$. Therefore, since at every iteration all the $8KN$ parameters of $\boldsymbol{\theta}$ are simultaneously updated, LS calibration has a very low speed of convergence. Furthermore, estimating a large number of unknowns, the Jacobian computation also becomes considerably costly.

3.2 SAGE calibration

In the case of solving for multiple sources in the sky, the SAGE calibration algorithm (Yatawatta et al. 2009; Kazemi et al. 2011) has a significantly improved computational cost and convergence rate compared to the LS calibration. The key point is that, in general, the SAGE algorithm (Fessler & Hero 1994) partitions the complete data space to smaller 'hidden' data spaces and estimates parameters in them rather than in the complete data space. Applying the SAGE algorithm to the calibration problem, the contribution of every lth source in the observation is assumed to depend only on a subset of parameters, $\boldsymbol{\theta}_l \in \mathbb{R}^{8N\times 1}$. Therefore, the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{8KN\times 1}$ could be partitioned for different directions (sources) in the sky as

$$\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1^{\mathrm{T}}\ \boldsymbol{\theta}_2^{\mathrm{T}} \cdots \boldsymbol{\theta}_K^{\mathrm{T}}\right]^{\mathrm{T}}.$$

This partitioning is justifiable when the sources are sufficiently separated from each other. Initializing a starting parameter vector $\widehat{\boldsymbol{\theta}}^1$, where $\widehat{\boldsymbol{\theta}}^k$ denotes the estimate of $\boldsymbol{\theta}$ obtained at the kth iteration, SAGE calibration algorithm is executed as follows:
for every iteration $k = 1, 2, \ldots$ until an upper limit for k or convergence of $\widehat{\boldsymbol{\theta}}^k$:

for all or some $l \in \{1, 2, \ldots, K\}$, update the lth source parameters $\boldsymbol{\theta}_l \in \mathbb{R}^{8N\times 1}$:

(i) Define the hidden data space as

$$\boldsymbol{x}_l = \boldsymbol{s}_l(\boldsymbol{\theta}_l) + \boldsymbol{n} \in \mathbb{R}^{4\tau N(N-1)\times 1}. \tag{7}$$

Thus, the observed data $\boldsymbol{y} \in \mathbb{R}^{4\tau N(N-1)\times 1}$ is given by

$$\boldsymbol{y} = \boldsymbol{x}_l + \sum_{\substack{z=1\\ z\neq l}}^{K} \boldsymbol{s}_z(\boldsymbol{\theta}_z). \tag{8}$$

(ii) *SAGE E step*: calculate the conditional mean $\widehat{\boldsymbol{x}}_l^k = \mathrm{E}\{\boldsymbol{x}_l|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}^k\}$ as

$$\widehat{\boldsymbol{x}}_l^k = \boldsymbol{s}_l\left(\widehat{\boldsymbol{\theta}}_l^k\right) + \left(\boldsymbol{y} - \sum_{z=1}^{K} \boldsymbol{s}_z\left(\widehat{\boldsymbol{\theta}}_z^k\right)\right) = \boldsymbol{y} - \sum_{\substack{z=1\\ z\neq l}}^{K} \boldsymbol{s}_z\left(\widehat{\boldsymbol{\theta}}_z^k\right).$$

(iii) *SAGE M step*: estimate

$$\widehat{\boldsymbol{\theta}}_l^{k+1} = \arg\min_{\boldsymbol{\theta}_l} ||[\widehat{\boldsymbol{x}}_l^k - \boldsymbol{s}_l(\boldsymbol{\theta}_l)]||^2,$$

by the LM method as

$$\widehat{\boldsymbol{\theta}}_l^{k+1} = \widehat{\boldsymbol{\theta}}_l^k - (\boldsymbol{\nabla}_{\boldsymbol{\theta}_l}^{\mathrm{T}}\boldsymbol{\nabla}_{\boldsymbol{\theta}_l} + \lambda\mathbf{H})^{-1}\boldsymbol{\nabla}_{\boldsymbol{\theta}_l}^{\mathrm{T}}\phi(\boldsymbol{\theta}_l)|_{\widehat{\boldsymbol{\theta}}_l^k}, \tag{9}$$

where $\phi(\boldsymbol{\theta}_l) = [\widehat{\boldsymbol{x}}_l^k - \boldsymbol{s}_l(\boldsymbol{\theta}_l)] \in \mathbb{R}^{4\tau N(N-1)\times 1}$.

endfor

endfor

Based on the above, at every kth iteration, SAGE method alternates between updating parameters of some or all the sources, $l \in \{1, 2, \ldots, K\}$. Calculating the ML estimate of $\boldsymbol{\theta}_l \in \mathbb{R}^{8N\times 1}$ in (9), instead of the ML estimate of all parameters $\boldsymbol{\theta} \in \mathbb{R}^{8KN\times 1}$ as in (6), it has been proved that the SAGE algorithm benefits from an accelerated convergence rate (Fessler & Hero 1994) compared to the LS method. The sizes of the Jacobian $\boldsymbol{\nabla}_{\boldsymbol{\theta}_l}$ and the linear system solved in (9) are $4\tau N(N-1)\times 8N$ and $8N$, respectively. In addition, the cost of computing $\boldsymbol{\nabla}_{\boldsymbol{\theta}_l}^{\mathrm{T}}\boldsymbol{\nabla}_{\boldsymbol{\theta}_l}$ is $\mathcal{O}((8N)^2 \times 4\tau N(N-1))$. Thus, applying LM algorithm for estimating $\boldsymbol{\theta}_l$ from (9), the computational expense of the SAGE calibration is much cheaper compared to the LS calibration.

Note that in the SAGE calibration, instead of partitioning the parameters of the individual sources, one could also make partitions including more than a single source sharing common parameters (Kazemi, Yatawatta & Zaroubi 2013). This is more efficient when some sources have a small angular separation from each other in the sky and hence share some parameters.

4 THE OS–LS AND OS–SAGE CALIBRATION METHODS

In this section, OS–LS (Liu, Yin & Shan 2005) and OS–SAGE (Hongqing et al. 2004) calibration algorithms, combinations of OS algorithm with LS and SAGE calibration methods, are introduced to speed up the conventional LS and SAGE calibration procedures.

OS algorithm is applied to those optimization problems with a cost function that can be expressed as a sum of several other cost functions for accelerating the convergence rate. The solutions obtained by the OS method attain almost the same accuracy as those obtained by the non-OS optimization methods in a fraction of the time (Hudson & Larkin 1994). The key idea is to consider the Jacobian of one, or some, subcost functions as an approximate gradient

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

of the original cost function. These approximations are quite reasonable when one is far from the optimal point, and provide OS method with a very fast convergence rate. However, at later iterations and when the parameters are close to the global optimum, the approximations restrict the OS method to a sub-optimal limit cycle (the optima of the individual sub-observations which are processed in OS iterations). Therefore, the OS method does not converge globally (Ahn 2004).

Denote the visibility vectors of the τ time and frequency samples that have the fixed gain errors $\boldsymbol{\theta} \in \mathbb{R}^{8KN\times 1}$ by $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_\tau$, where $\boldsymbol{y}_t \in \mathbb{R}^{4N(N-1)\times 1}$ for $t \in \{1, 2, \ldots, \tau\}$. Since the noise is statistically independent, calibration problem could be restated as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \prod_{t=1}^{\tau} f_t(\boldsymbol{y}_t; \boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{\tau} \mathcal{L}_t(\boldsymbol{\theta}|\boldsymbol{y}_t), \tag{10}$$

where f_t and $\mathcal{L}_t$ are the probability density and the log-likelihood functions for the visibility vector $\boldsymbol{y}_t$, respectively. OS algorithm is applied for accelerating the maximization of this sum of log-likelihood functions. Supposing that the following Jacobian equivalence conditions hold

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_1 \cong \nabla_{\boldsymbol{\theta}} \mathcal{L}_2 \cong \cdots \cong \nabla_{\boldsymbol{\theta}} \mathcal{L}_\tau, \tag{11}$$

then the OS method sequentially updates the parameters $\boldsymbol{\theta}$ for one or some visibility vectors $\boldsymbol{y}_t$ (sub-observations). The solution of every sub-observation is used as the starting point of the next sub-observation. Since each subcost function $\mathcal{L}_t$ involves a subset of data, $\boldsymbol{y}_t$, which is independent from the others, the method is named 'ordered subsets'. Sub-observations might be ordered for updating by some scheme that gives preferences to the data items, or, as in this work, in random. An introduction to the OS algorithm is presented by Ahn (2004). In the following, the OS–LS and OS–SAGE methods are outlined. Note that the size of sub-observations $\boldsymbol{y}_t$s must be greater than or equal to the number of unknown parameters in $\boldsymbol{\theta}$.

4.1 OS–LS calibration

In the presented OS–LS calibration, the LM method is selected as the gradient-based ML estimation algorithm of the LS calibration. Starting with an initial suggestion $\widehat{\boldsymbol{\theta}}^1 \in \mathbb{R}^{8KN\times 1}$, OS–LS is executed as

for every iteration $k = 1, 2, \ldots$ until an upper limit or convergence of $\widehat{\boldsymbol{\theta}}^k$, run *m OS iterations*:

for some or all sub-observation $\{\boldsymbol{y}_t | t = 1, \ldots, m \leq \tau\}$:
Select $\boldsymbol{\theta}^k = \widehat{\boldsymbol{\theta}}^t$, and calculate

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - (\nabla_{\boldsymbol{\theta}}^{\mathrm{T}} \nabla_{\boldsymbol{\theta}} + \lambda \mathbf{H})^{-1} \nabla_{\boldsymbol{\theta}}^{\mathrm{T}} \phi(\boldsymbol{\theta})|_{\boldsymbol{\theta}^k}, \tag{12}$$

where $\phi(\boldsymbol{\theta}) = [\boldsymbol{y}_t - \sum_{l=1}^{K} \boldsymbol{s}_l(\boldsymbol{\theta})] \in \mathbb{R}^{4N(N-1)\times 1}$.
Select $\widehat{\boldsymbol{\theta}}^{(t \bmod m)+1} = \boldsymbol{\theta}^{k+1}$ for the next sub-observation.
endfor

endfor

As given above, at every LM iteration, parameters are sequentially updated for some or all sub-observations. The sizes of the Jacobian $\nabla_{\boldsymbol{\theta}}$ and the linear system solved in (12) are $4N(N-1) \times 8KN$ and $8KN$, respectively. Moreover, the cost of computing $\nabla_{\boldsymbol{\theta}}^{\mathrm{T}} \nabla_{\boldsymbol{\theta}}$ is $\mathcal{O}((8KN)^2 \times 4\,N(N-1))$. When (11) holds, the Jacobian is calculated only for one, or a few, number of sub-observations per iteration and hence, the OS–LS method's convergence rate is considerably increased compared to the LS method.

4.2 OS–SAGE calibration

In this section, the OS–SAGE calibration method is introduced. A similar OS–SAGE technique is used for positron emission tomography (PET) by Hongqing et al. (2004).

Initializing $\widehat{\boldsymbol{\theta}}^1 \in \mathbb{R}^{8N\times 1}$, OS–SAGE is outlined as follows:
for every $k = 1, 2, \ldots$ until an upper limit for k or convergence of $\widehat{\boldsymbol{\theta}}^k$, execute *m OS iterations*:

for some or all sub-observations $\{\boldsymbol{y}_t | t = 1, \ldots, m \leq \tau\}$:
Select $\boldsymbol{\theta}^k = \widehat{\boldsymbol{\theta}}^t$.

for all or some $l \in \{1, 2, \ldots, K\}$, update the lth source parameters $\boldsymbol{\theta}_l \in \mathbb{R}^{8N\times 1}$:

(i) Define

$$\boldsymbol{y}_t = \boldsymbol{x}_l + \sum_{\substack{z=1 \\ z\neq 1}}^{K} \boldsymbol{s}_z(\boldsymbol{\theta}_z), \quad \boldsymbol{x}_l = \boldsymbol{s}_l(\boldsymbol{\theta}_l) + \boldsymbol{n}.$$

(ii) *SAGE E step*: calculate $\widehat{\boldsymbol{x}}_l^k = \mathrm{E}\{\boldsymbol{x}_l | \boldsymbol{y}_t, \boldsymbol{\theta}^k\}$ as

$$\widehat{\boldsymbol{x}}_l^k = \boldsymbol{y}_t - \sum_{\substack{z=1 \\ z\neq 1}}^{K} \boldsymbol{s}_z(\boldsymbol{\theta}_z^k), \ \boldsymbol{y}_t \in \mathbb{R}^{4N(N-1)\times 1}.$$

(iii) *SAGE M step*: similar to (9), estimate

$$\boldsymbol{\theta}_l^{k+1} = \arg\min_{\boldsymbol{\theta}_l} ||[\widehat{\boldsymbol{x}}_l^k - \boldsymbol{s}_l(\boldsymbol{\theta}_l)]||^2,$$

using the LM method, by

$$\boldsymbol{\theta}_l^{k+1} = \boldsymbol{\theta}_l^k - \left(\nabla_{\boldsymbol{\theta}_l}^{\mathrm{T}} \nabla_{\boldsymbol{\theta}_l} + \lambda \mathbf{H}\right)^{-1} \nabla_{\boldsymbol{\theta}_l}^{\mathrm{T}} \phi(\boldsymbol{\theta}_l)|_{\boldsymbol{\theta}_l^k} \tag{13}$$

endfor
Select $\widehat{\boldsymbol{\theta}}^{(t \bmod m)+1} = \boldsymbol{\theta}^{k+1}$ for the next sub-observation.
endfor

endfor

OS method reduces the data size from $4\tau N(N-1)$ to $4N(N-1)$, since it calculates the partial gradients for sub-observations $\boldsymbol{y}_t \in \mathbb{R}^{4N(N-1)\times 1}$, $t \in \{1, 2, \ldots, \tau\}$, instead of the whole observed data $\boldsymbol{y} \in \mathbb{R}^{4\tau N(N-1)\times 1}$. Thus, the size of the Jacobian $\nabla_{\boldsymbol{\theta}_l}$, where $\phi(\boldsymbol{\theta}_l) = [\widehat{\boldsymbol{x}}_l^k - \boldsymbol{s}_l(\boldsymbol{\theta}_l)] \in \mathbb{R}^{4N(N-1)\times 1}$, calculated by LM method for every OS iteration of the OS–SAGE calibration at (13), is $4N(N-1) \times 8N$. The size of the linear system solved in (13) is $8N$ and the cost of computing $\nabla_{\boldsymbol{\theta}_l}^{\mathrm{T}} \nabla_{\boldsymbol{\theta}_l}$ is $\mathcal{O}((8N)^2 \times 4\,N(N-1))$. When $m \ll \tau$, the OS–SAGE method converges much faster than the conventional SAGE algorithm for which the Jacobian size is $4\tau N(N-1) \times 8N$. On the other hand, for every tth OS iteration, the updated result of the $(t-1)$th sub-observation is used as the starting point. Every OS–SAGE iteration includes m number of SAGE iterations. Therefore, at initial iterations when (11) holds, OS–SAGE algorithm increases the likelihood function as equivalent to SAGE method with m iterations. Thus, the convergence of OS–SAGE compared with SAGE is accelerated.

4.3 Partitioning the baselines

So far, we have divided the data into sub-observations only based on their integration time and frequency. However, there are cases in

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

which we need to calibrate for a single time and frequency interval. For instance, consider calibrating only for the ith time and frequency interval when $1 \leq i \leq \tau$. To apply OS calibration to such a case, one can define sub-observations by partitioning the data vector $\boldsymbol{y}_i$ over the instrument's baselines as

$$\boldsymbol{y}_i = \left[\boldsymbol{y}_{i1}^{\mathrm{T}}\ \boldsymbol{y}_{i2}^{\mathrm{T}}\ \cdots\ \boldsymbol{y}_{iB}^{\mathrm{T}}\right]^{\mathrm{T}}, \quad B \ll \frac{N(N-1)}{2}.$$

Then, similar to (11), the calibration problem becomes

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{b=1}^{B} \mathcal{L}_b(\boldsymbol{\theta}|\boldsymbol{y}_{ib}) \tag{14}$$

for which OS methods presented by Sections 4.1 and 4.2 are applicable, and where *OS iterations* are executed over $\{\boldsymbol{y}_{ib}|b = 1, \ldots, m \leq B\}$. Utilizing such an OS calibration could also be beneficial in cutting down the computational expense of calibration of interferometers with a large number of receivers. The only points that should be taken into account are the following.

(i) Every partition of data (sub-observation) $\boldsymbol{y}_{ib}$, for $b \in \{1, 2, \ldots, B\}$, must have visibilities from different baselines such that the baselines cover all the receivers of the instrument (or all the parameters).

(ii) The number of visibilities of every sub-observation must be equal to, or larger than, the number of calibration unknowns,

$$||\boldsymbol{y}_{ib}||_1 \geq 8KN. \tag{15}$$

4.4 Discussion

To wrap up all the discussed calibration algorithms, we present a general overview in Fig. 1. Fig. 1 illustrates LS, SAGE, OS–LS and OS–SAGE calibrations algorithms.

Note that the following.

(i) As it is discussed at the beginning of Section 4, the OS algorithms do not necessarily converge. Nevertheless, there exist two major approaches in dealing with the convergence problem of the OS method: (a) using relaxation parameters (step sizes) (Ahn 2004). Calculating suitable relaxation parameters per every iteration is considerably costly. That makes the approach of progressively decreasing the number of sub-observations in OS method to be preferable. (b) Reducing the number of subsets with increasing iterations until the complete data set estimate is reconstructed (Hudson & Larkin 1994). In the OS method, one can incrementally combine some sub-observations together until there are no individual sub-observations remaining. Therefore, at the final iteration, the OS method is in fact changed to the non-OS technique which is used for the ML approximations, solving for the complete data set. This approach guarantees global convergence as long as the non-OS ML estimation techniques (LS, SAGE, etc.) converge. However, it must be taken into account that the highest accuracy achievable by the proposed scheme is equal to any non-OS optimization methods. Modifying OS calibration in order to achieve an accuracy superior to the ones obtained by non-OS calibrations is addressed in future work.

(ii) When the signal-to-noise ratio (SNR) is poor, shifting to non-OS calibrations after running a few number of OS iterations is recommended. Moreover, instead of running the OS method on every individual time and frequency sub-observation $\boldsymbol{y}_t$, for

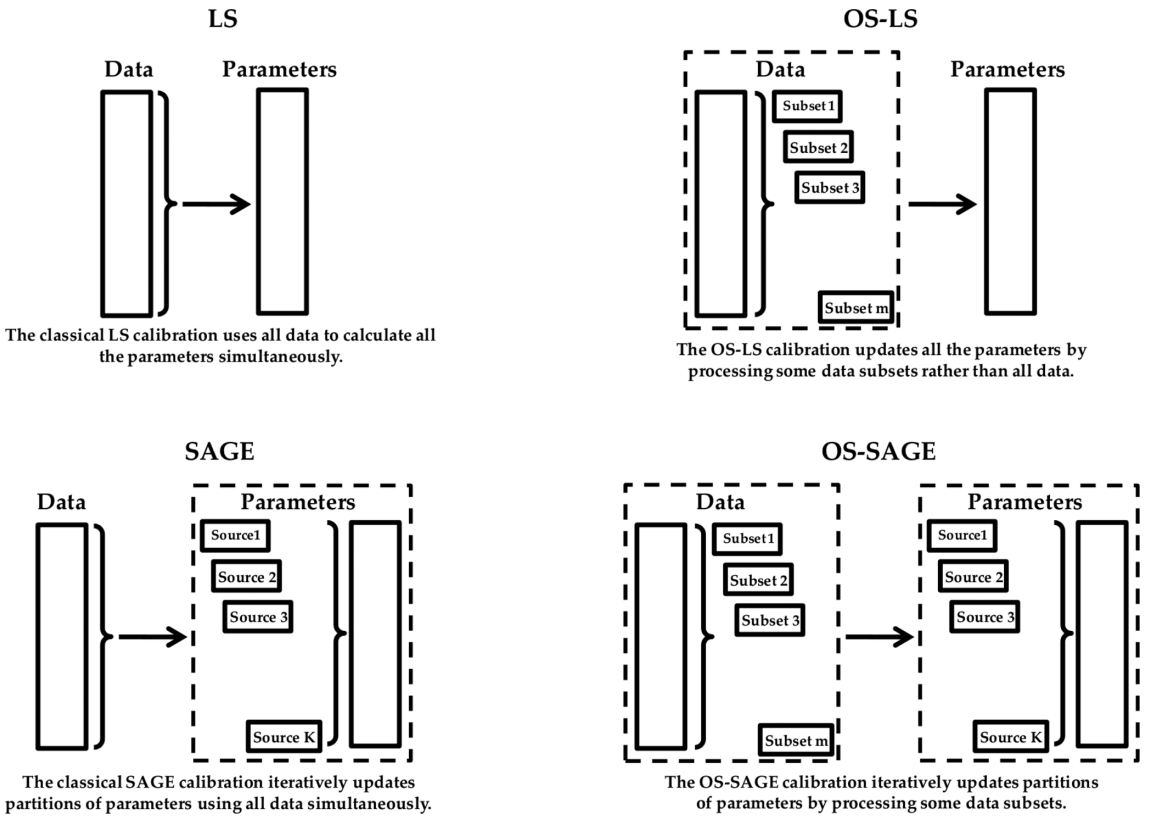

Figure 1. Diagrams illustrating LS, SAGE, OS–LS and OS–SAGE calibrations algorithms.

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

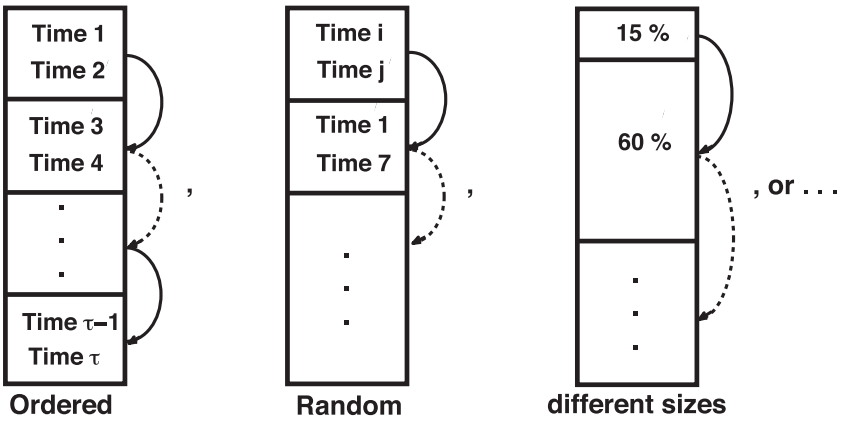

Figure 2. Instead of running the OS method on every individual sub-observation, one could also apply the method to combinations of two or more sub-observations to improve the SNR. Examples of having incrementally ordered data sets of size two, randomly chosen data sets of the same size and randomly chosen data sets from different sizes are shown from left to right, respectively.

$t \in \{1, 2, \ldots, \tau\}$, one could also apply the method to combinations of two or more sub-observations to improve the SNR. Fig. 2 shows examples of having incrementally ordered data sets of size two, randomly chosen data sets of the same size and randomly chosen data sets from different sizes, from left to right, respectively. The data sets could be arranged in different orders depending on the characteristics of specific observations. Similarly, subsets of frequency ordered sub-observations could be introduced.

(iii) In the calibration data model presented by (4), we consider a very general form of the Jones matrices $\mathbf{J}$, as complex 2×2 matrices, and then search for the real and imaginary parts of their elements which are collected in $\boldsymbol{\theta}$. However, one can use a more detailed presentation of the Jones matrices in the data model, for instance, when the elements of the Jones matrices are functions of time ζ and frequency ξ,

$$\mathbf{J} = \begin{bmatrix} \eta_1(\zeta, \xi) & \eta_2(\zeta, \xi) \\ \eta_3(\zeta, \xi) & \eta_4(\zeta, \xi) \end{bmatrix}. \tag{16}$$

Then, calibration is estimation of these functions, denoted by η in (16). But, this leads again to estimation of some constant parameters which define the functions. Therefore, OS calibration is also useful for such a case as well and its partitioning of data to time and frequency sub-observations would not cause any degradation of the accuracy of calibration.

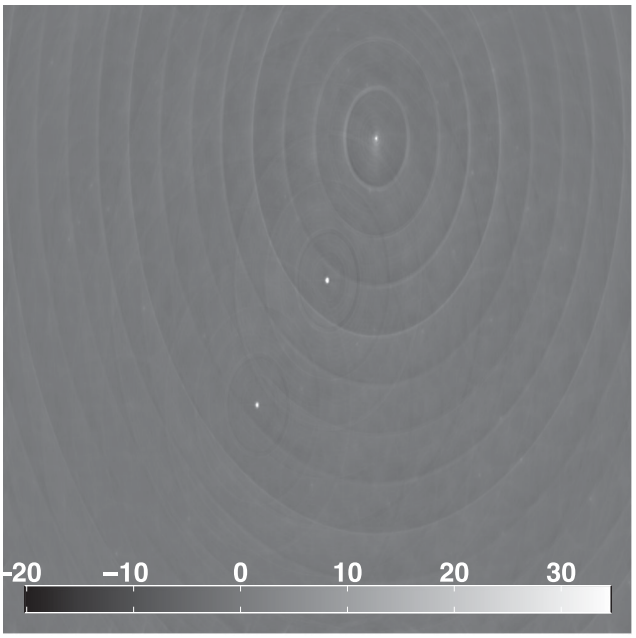

Figure 3. An $8^\circ \times 8^\circ$ WSRT 12 h simulated observation of three bright sources, with intensities 160, 107 and 108 Jy, and 47 faint sources, with intensities below 15 Jy. The frequency is 355 MHz and the SNR is 80.

5 RESULTS

In this section, simulated data are used to compare the performance of LS and SAGE calibrations with OS–LS and OS–SAGE ones. Note that n in this section denotes the number of iterations of the conventional LS and SAGE methods. The implementation of the calibration algorithms are done using MATLAB software. The unit of colour bars of all the images are in Jansky (Jy).

5.1 Simulations

A 12-h observation of Westerbork Synthesis Radio Telescope (WSRT), including 14 receivers observing a sky with 50 sources, is simulated. Three sources are very bright with intensities 160, 107 and 108 Jy, and 47 other sources are faint with intensities below 15 Jy. The source positions are following a uniform distribution. The Jones matrices are generated as multiplications of different linear combinations of sin and cos functions. Their gradients vary slowly (coherence time about 3 min) as a function of time such that on a few seconds time intervals the variation could be negligible. We keep the SNR = 80. The simulated single channel image at 355 MHz is shown in Fig. 3 in which the background faint sources are almost invisible.

We partition the simulated data to 10 s time intervals, $\tau = 10$, including sub-observations obtained from 10 individual seconds, for which the gain errors are assumed to be the same. Then, we calibrate the data partitions only for the three brightest sources via the LS and SAGE calibration methods. The residual images, obtained after $n = 9$ iterations, are presented in Fig. 4. As Fig. 4 shows, among those three subtracted bright sources, the central one is the best removed (slightly underestimated) by both SAGE and the LS calibration methods. The unsolved 47 faint sources are also visible in both residual images. But, the two other bright sources are not

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

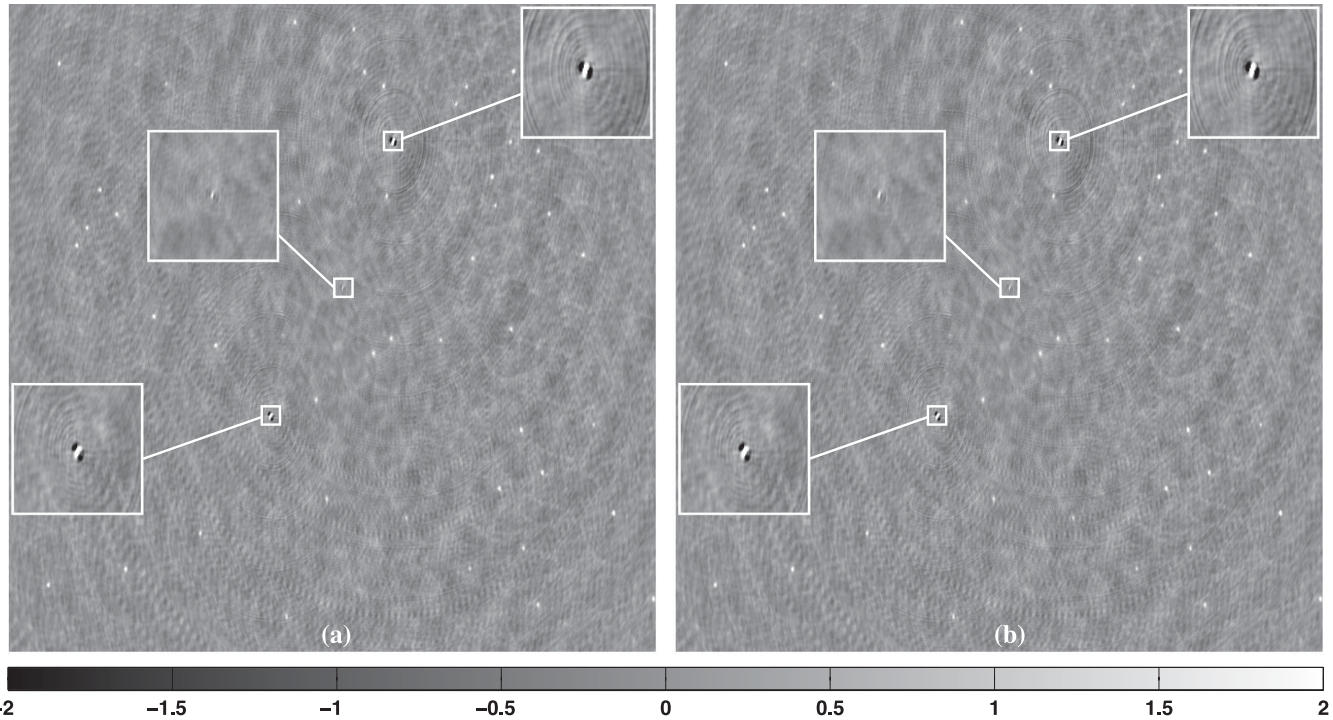

Figure 4. The residual images of the LS (a) and the SAGE (b) calibrations, solving only for the three brightest sources with $n = 9$ iterations. Calibrations are executed on every $\tau = 10$ sub-observations simultaneously. From the three subtracted sources, the central one is the best removed (slightly underestimated) by both the SAGE and the LS calibrations. The unsolved 47 faint sources are also visible in both (a) and (b). However, the two other bright sources are not subtracted perfectly which is expected to be improved by increasing the number of iterations. The residual noise of the SAGE algorithm is lower than of the LS method (Table 1). This reveals the superior performance of the SAGE calibration compared to the LS calibration.

Table 1. Execution times of calibration (min) and the standard deviation of the residual noise (mJy). The OS calibrations perform much faster than the conventional LS and SAGE calibrations. The lowest execution time of the OS results is obtained for $m = 1$. On the other hand, the most accurate results are obtained for $m = 2$ number of OS iterations. Moreover, SAGE-type calibrations are always preferred to the LS ones, having a higher accuracy and less computational complexity.

m = number of OS iterations	LS		SAGE	
LS or SAGE iterations $n = 9$	Time (min)	Noise (mJy)	Time (min)	Noise (mJy)
OS, $m = 1$	41.3	234.2	9.7	226.1
OS, $m = 2$	75.5	232.9	20.4	225.7
Conventional methods	103.9	180.1	86.3	179.2

subtracted perfectly (overestimated in the left- and right-hand sides and underestimated in the central parts). This problematic pattern is expected to be improved by increasing the number of iterations. There is no significant difference between the residual images produced by the LS and SAGE methods in Fig. 4. However, as it is shown in Table 1, the noise level in the residual image of the SAGE calibration is lower than the one of the LS method. Therefore, SAGE calibration reveals a superior performance compared to the LS calibration since it achieves more accurate results with a considerably less computational complexity (Kazemi et al. 2011).

The data are also calibrated by the OS–LS and OS–SAGE methods using $n = 9$ iterations. OS iterations are executed for $m = 1$, 2 number of sub-observations which are randomly chosen. The residual images after subtracting the three brightest sources are presented in Fig. 5. As Fig. 5 shows, the central source becomes problematic in the results of the OS calibrations and it was much better removed by the conventional LS and SAGE calibrations in Fig. 4. Except for this source, the OS calibrations have a similar quality in the residual images to the conventional LS and SAGE calibrations. The two other subtracted sources are not perfectly removed and the other 47 faint sources are visible in the images, similar to Fig. 4. The residual images obtained for $m = 1$ and $m = 2$ OS iterations look almost the same. There is no significant improvement in the residual noise level when using $m = 2$ OS iterations instead of $m = 1$, as it is evident in Table 1. In this case, the OS calibration with $m = 1$ OS iteration is preferable in comparison with $m = 2$ since it carries a lower computational cost.

The calibrations execution times, in minutes, and the residual noise levels, in milliJansky (mJy), are presented in Table 1. Table 1 shows that the OS calibrations have a much faster processing speed compared to the conventional LS and SAGE calibrations. Among OS calibrations, the ones with a smaller number of OS iterations always have faster execution, as it is the case comparing the processing times for $m = 1$ to $m = 2$. The fastest execution speed of the calibration method belongs to the OS calibrations with $m = 1$ OS iteration. On the other hand, the OS calibrations including a large number of OS iterations usually produce more accurate solutions since they use a higher level of information in their computations. As the results of Table 1 demonstrate, the accuracy obtained by $m = 2$ number of OS iterations is slightly higher than the one achieved by $m = 1$. However, the use of $m = 1$ number of OS iterations is still preferred compared to $m = 2$ since it has a considerably lower processing time. Note that the use of the SAGE-type calibration methods is always preferred compared to the LS ones, providing more accurate results in a lower processing time.

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

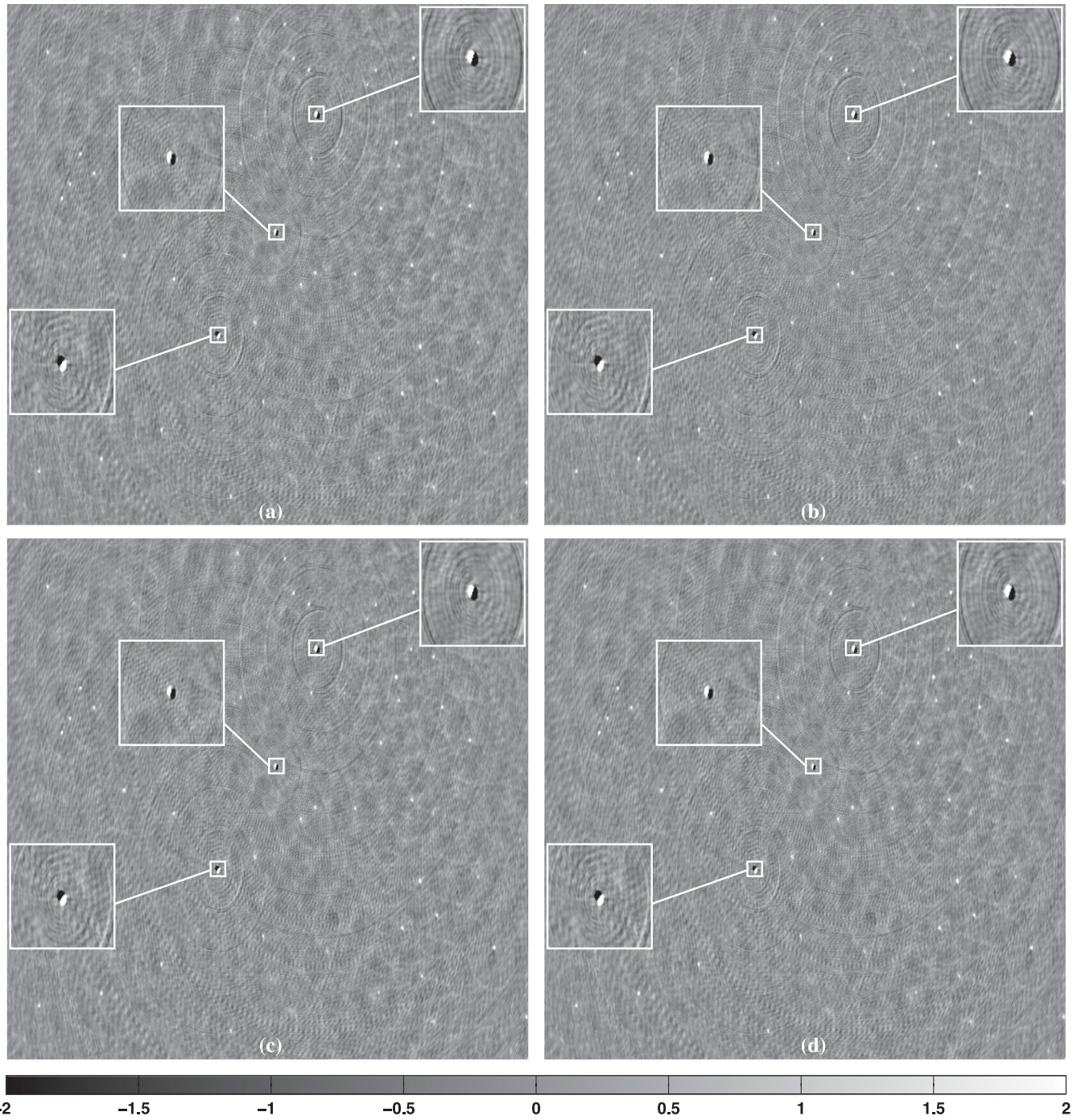

Figure 5. The residual images of the OS–LS calibration for $m = 1$ (a) and $m = 2$ (c), and the OS–SAGE calibration for $m = 1$ (b) and $m = 2$ (d) OS iterations. Calibrations are executed for the three brightest sources using $n = 9$ iterations. The central source is problematic in the residuals of the OS calibrations and was much better removed in the results of the conventional LS and SAGE calibrations presented in Fig. 4. Except for this source, the residual images obtained by the OS calibrations maintain the quality of the ones produced by the conventional LS and SAGE calibrations in Fig. 4. There is no visible difference between the results of $m = 1$ and $m = 2$ OS iterations in the images. That makes the OS calibration with $m = 1$ OS iteration preferable in comparison with $m = 2$ since it carries a lower computational cost.

Fig. 6 illustrates the residual noise level achieved by the calibration procedures versus the number of iterations of the LS and SAGE methods, when it varies between one and nine, $n \in \{1, \ldots, 9\}$. The number of OS iterations is denoted by m. In the plots of Fig. 6, the residual noise levels of the OS calibrations are higher than the ones of the non-OS calibrations. However, it must be taken into account that these results are obtained by using a comparably less computational cost compared to the classical LS and SAGE calibrations. By increasing n, the result of SAGE calibrations is always better than the one of LS calibrations. Moreover, the accuracy of OS calibration using $m = 2$ OS iterations is also always superior to the results obtained by $m = 1$.

As we have seen so far in this simulation, among the OS calibrations, the ones with a smaller number of OS iterations (smaller m)

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

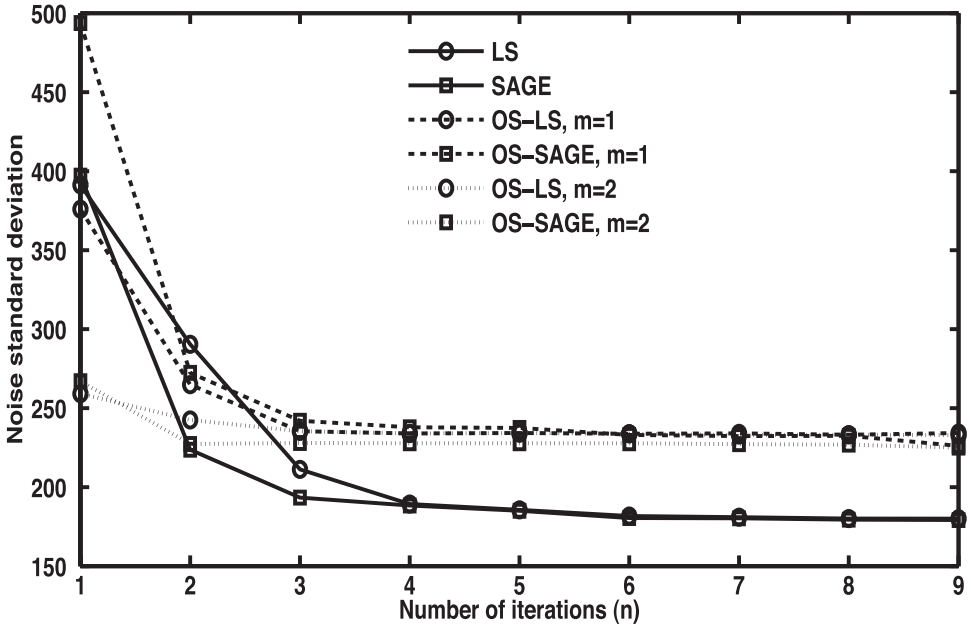

Figure 6. The residual noise standard deviations of the calibration methods in (mJy) versus their number of iterations which varies between one and nine, $n \in \{1, \ldots, 9\}$. The number of OS iterations are denoted by m. In the plots of Fig. 6, the residual noise levels of the OS calibrations are higher than the ones obtained by the non-OS calibrations. However, it must be taken into account that these results are generated spending a comparably less computational cost compared to the classical LS and SAGE calibrations. By increasing n, the result of SAGE calibrations are always better than of LS calibrations. Moreover, the accuracy of OS calibrations which use $m = 2$ OS iterations is superior to the one obtained by $m = 1$.

have a lower execution time. On the other hand, the OS calibrations including a large number of OS iterations usually produce more accurate solutions since they use a higher level of information in their computations of the Jacobian. However, the use of a small number of OS iterations is still preferable since it is considerably faster and applying a high enough number of calibration iterations, we would achieve the same accuracy as with large m.

In this section, we also demonstrate the applicability of the OS calibration in calibrating for a single time and frequency data sample, as it is discussed in Section 4.3, where the data must be partitioned over the instrument's baselines. There are various ways of such a partitioning of visibilities among which we use the most efficient one for this specific simulation.

(i) The first question is '*what is the maximum number of partitions of data over the baselines that we can define such that the baselines of every single partition cover all the receivers of the interferometer?*'. The reason of searching the maximum is to get the highest level of information at every calibration's sub-observation later on. To answer this question, we use some well-known definitions of graph theory (Diestel 2012).

Consider the interferometer as a complete graph of order N[8] where the receivers and the baselines are the nodes and edges of the graph, respectively. Therefore, since in this simulation N is even, the answer to our question is the chromatic index of this graph which is equal to $N - 1$. This means we can colour the $\frac{N(N-1)}{2}$ edges of the graph by $N - 1$ colours where every colour is covering all the N nodes and $\frac{N}{2}$ number of edges. For instance, Fig. 7 shows a complete graph of order eight, coloured by $8 - 1 = 7$ colours, where every colour covers all the nodes by $\frac{8}{2} = 4$ number of edges. We partition the

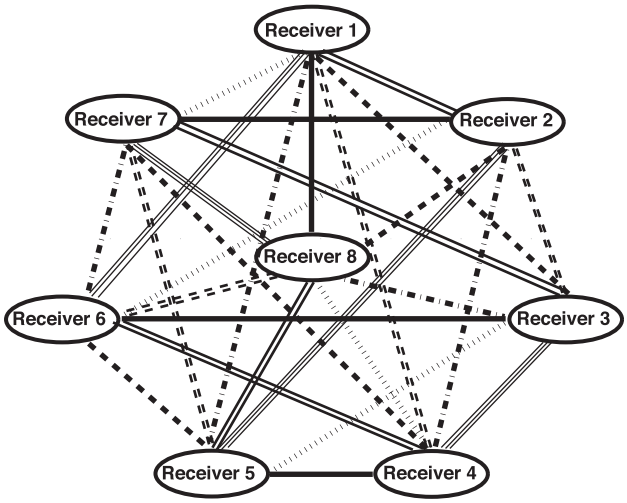

Figure 7. A complete graph of order eight, coloured by $8 - 1 = 7$ number of colours. Every colour covers all the nodes by $\frac{8}{2} = 4$ number of edges.

visibilities based on the colour of their corresponding baselines in the graph. Thus, at every partition, we have $\frac{N}{2}$ number of visibility matrices.

(ii) The second question is '*how many partitions should be collected at every OS calibration's sub-observation to ensure that (4) is not an underdetermined system?*'. Every partition has $\frac{N}{2}$ of baselines and we are trying to estimate KN Jones matrices. Therefore, we must have at least x partitions at every OS calibration's sub-data set where

$$x\frac{N}{2} > KN. \tag{17}$$

[8] A complete graph of order N has N nodes and every pair of nodes is connected to each other by a unique edge.

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

Thus,

$$x \geq 2K + 1. \tag{18}$$

We have $N = 14$ number of receivers in WSRT. Thus, $\frac{N(N-1)}{2} = 91$ number of baselines, providing 2×2 visibility matrices, at every time and frequency sample. According to (i) we can make 13 partitions of baselines so that every partition includes $\frac{N}{2} = 7$ number of visibilities covering all the receivers. Since we calibrate for $K = 3$ bright sources A, B and C, using (ii), $x \geq 7$. This means at every OS sub-observation we must collect at least seven numbers of those partitions. Thus, at every sub-observation we have $x \times \frac{N}{2} = 49$ number of visibility matrices and that is enough for estimating $KN = 42$ number of Jones matrices. Indeed better accuracy of OS calibration is expected to be obtained by increasing x till $x \leq N - 1$. This approach of defining sub-observations of the OS calibration is demonstrated in Fig. 8. As this figure shows, there are no overlaps between the baselines of the 13 different partitions. Therefore, the maximum information level, achievable by using $x \times \frac{N}{2} = 49$ number of visibilities, is provided for every sub-observation of the OS calibration.

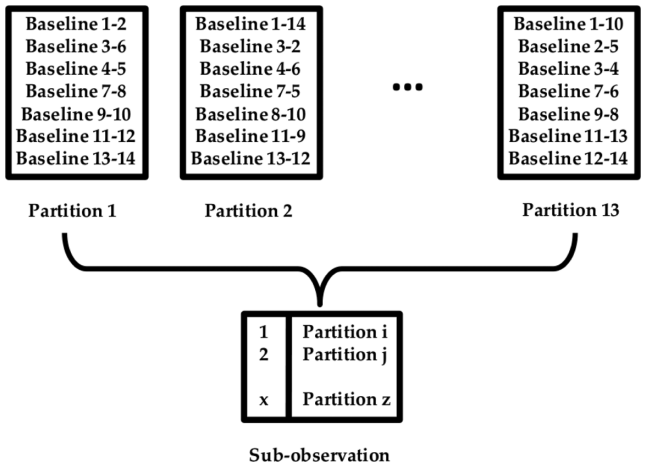

Figure 8. There are $N = 14$ number of receivers in WSRT and hence $N - 1 = 13$ number of partitions over its baselines, each including $\frac{N}{2} = 7$ number of visibilities and covering all the receivers. Every OS sub-observation consists of $x \leq 7$ number of such partitions. Thus, at every sub-observation we have $x \times \frac{N}{2}$ number of visibility matrices.

OS–SAGE calibration is executed, using $m = 2$ number of time samples at every iteration (two number of OS iterations) for $x = 7$ and $x = 10$. The residual images are shown in Fig. 9. We can see that by increasing the number of visibilities in the sub-observations from 49 ($x = 7$) to 70 ($x = 10$), the calibration accuracy is highly improved. We also can see that the two images of Fig. 9 have a higher residual noise and artefacts compared to the result obtained for $x = N - 1 = 13$, which is presented in Fig. 5 as image (d). This shows that better accuracy of the OS calibration is achieved when the number of visibilities in every sub-observation is large. However, the calibration's processing times for $x = 7$ and $x = 10$ are 73.5 and 92.8 min, respectively, while for $x = 13$ it is 108.8 min (Table 1). Remember that the whole point of partitioning the baselines was to cut down the computations. We also can benefit from this approach to speed up the initial calibration iterations for the telescope with a large number of baselines such as SKA.

As a final remark, for partitioning baselines of a telescope with an odd number of receivers N, an alternative would be (i) first partitioning baselines for $N - 1$ number of receivers, as it is already

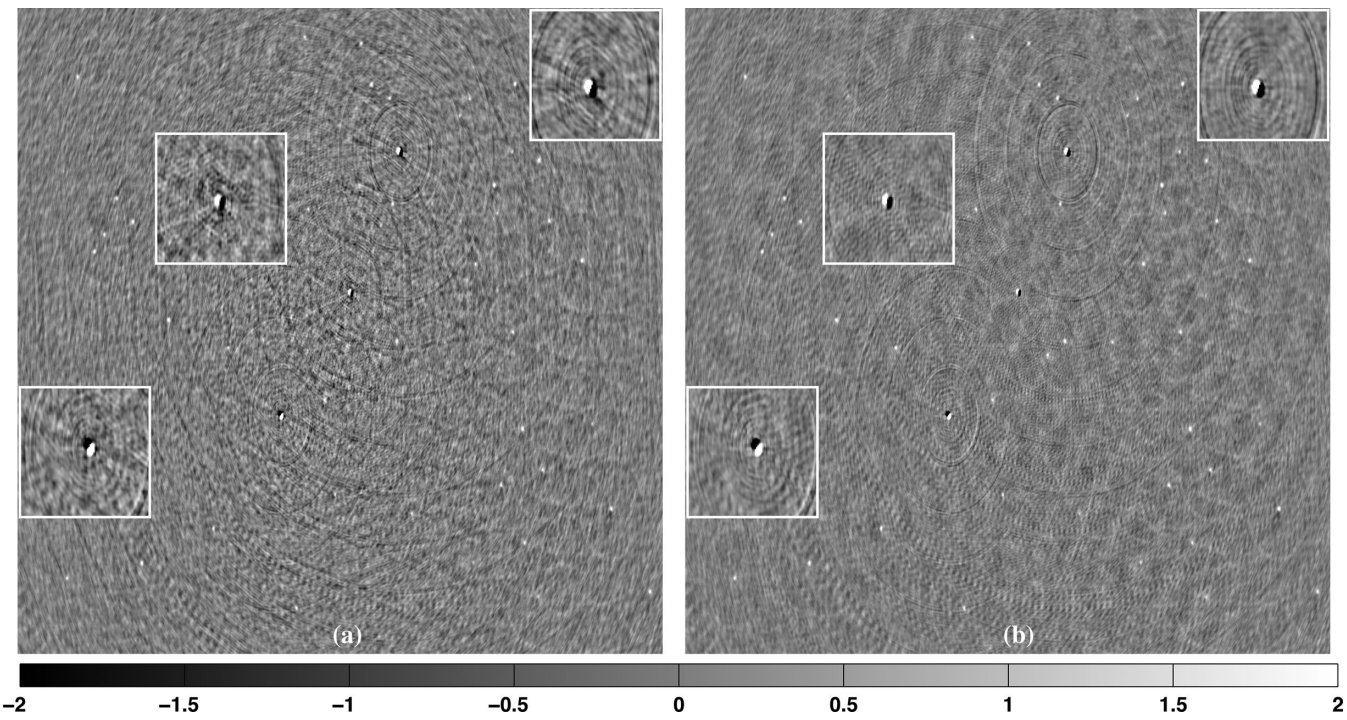

Figure 9. The residual images obtained by the OS–SAGE calibration, using $m = 2$ number of sub-observations at every iteration for $x = 7$ (a) and $x = 10$ (b). By increasing the number of visibilities in the sub-observations from 49 ($x = 7$) to 70 ($x = 10$), the calibration accuracy is highly improved. Plus, the two images have a higher residual noise and artefacts compared to the result obtained for $x = N - 1 = 13$, which is presented in Fig. 5 as image (d). This shows that better accuracy of the OS calibration is achieved when there exist a larger number of visibilities in every sub-observation. However, the calibration's processing time for $x = 7$ and $x = 10$ is 73.5 and 92.8 min, respectively, which is faster than the one for $x = 13$ that was 108.8 min (Table 1).

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

explained in this section, and (ii) assigning the remained baselines to these $N-1$ partitions.

5.2 Averaging of visibilities

The OS calibration method divides the data into sub-observations and alternates. The use of fewer data samples in each iteration is the principle cause of the speedup. So far, we have used segments of data consisting of multiple integrations in time and have considered the individual integrations as the sub-observations. This is reasonable for the use of OS calibrations. However, for the non-OS-type calibrations all of these integrations are explicitly considered to be equivalent. Therefore, one could ask if it is easier to average the data before calibration to decrease the computational cost.

To answer this question, consider the case of calibrating data for a point source far away from the phase centre of an observation. Based on (3), the visibilities of baseline $p-q$ at every sub-observation are formulated as

$$\boldsymbol{v}_{pq} = \mathbf{J}_q^* \otimes \mathbf{J}_p \mathrm{vec}(\mathbf{C}) + \boldsymbol{n}_{pq}, \tag{19}$$

where

$$\mathbf{C} = \mathrm{e}^{\frac{-2\pi j\xi}{c}(ul+vm+w(\sqrt{1-l^2-m^2}-1))} \begin{bmatrix} \frac{I}{2} & 0 \\ 0 & \frac{I}{2} \end{bmatrix}. \tag{20}$$

In (20), $j^2=-1$, ξ is the frequency of the observation, c is the speed of light, (l, m) are the source direction components corresponding to the observation phase centre, (u, v, w) are the geometric components of baseline $p-q$ and I is the intensity of the source.

Since the source is far away from the phase centre, (l, m) in (20) are large. Therefore, even very small variation of the baselines (u, v, w) on different sub-observations cause huge differences in the phase terms of (20). Subsequently, averaging the visibilities of (19) causes de-correlation (losing amplitude) and smearing effects in the calibration residuals.

To illustrate this, we simulate a 12-h observation of WSRT from a very bright source with 130 Jy intensity is simulated. The source is about 4° away from the phase centre. In the centre of the field we also put 23 faint sources with intensities below 9 Jy. The Jones matrices for the faint sources are considered as identity matrices. For the bright source, they are multiplications of different linear combinations of sin and cos functions which are invariant on 25 s time intervals. That provides time samples of size $\tau=25$ including sub-observations, from every individual second, for which the gain errors are exactly the same. White Gaussian noise is also added to the simulated data(see Fig. 10).

It is expected that traditional calibration after averaging data performs as equivalent as the OS calibration which iterates on the individual sub-observations. The reason is that the simulated corruptions in the signals on 25 s time intervals are invariant. However, the results, illustrated in Fig. 11, are completely the opposite.

Fig. 11 shows the residual images obtained by the LS and the OS–LS calibration, utilizing $m=2$ number of OS iterations and $n=9$ number of LS iterations. The processing time in minutes is shown at the bottom right-hand corner of every image. In image (a) of Fig. 11, LS calibration is applied on averaged data obtained from $\tau=25$ time samples. In this image, the bright source is highly underestimated (almost not subtracted at all) and there exist elongated radial features. This is due to the de-correlation by averaging the visibilities. However, in image (b) of Fig. 11, for which OS–LS calibration is applied on individual integrations, the bright source is perfectly subtracted and the other fainter sources are completely visible. This proves that we cannot simply apply calibration on averaged visibilities to cut down the computations and reveals the need of using the OS calibration. We have also executed LS calibration on non-averaged data sets of $\tau=25$ time samples. The resulted residual image has been exactly the same as image (b) of Fig. 11, which is generated by OS–LS calibration. The reason is that the Jones matrices on every 25 s calibrated data are invariant. Therefore, the solution which is obtained by OS calibrations, using few integrations (sub-observations) within 25 s, is the same as the one obtained by non-OS calibrations using all the data. However, in reality, Jones matrices vary with time. In such a case, the result of the non-OS calibrations is always better than, or equivalent to, the one of OS calibrations. It is because finding a global solution which fits all data is generally more efficient than solving only for a part of data set. The execution time of the LS calibration was 78.15 min, which is indeed longer than the one of OS–LS calibration (19.43 min).

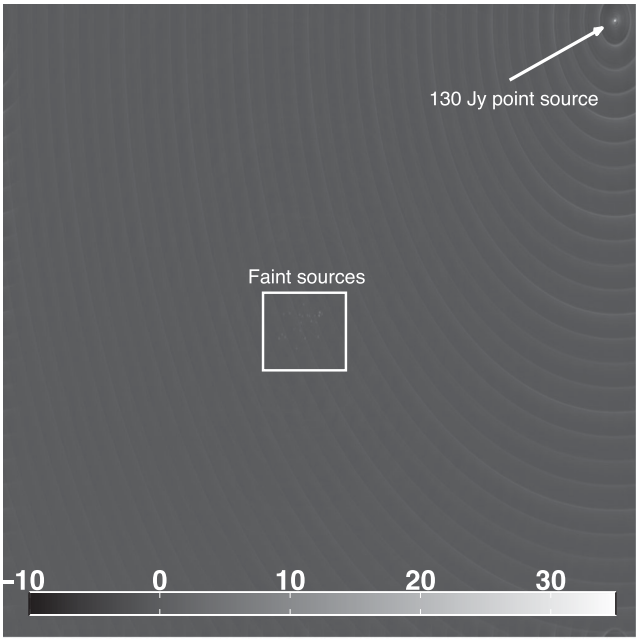

Figure 10. A 12-h simulation of WSRT with a bright, 130 Jy source. The source is about 4° away from the phase centre. There also exist 23 other faint sources with intensities below 9 Jy in the centre of the field. White Gaussian noise is also added to the simulated data.

6 CONCLUSIONS

This paper introduces OS–LS and OS–SAGE radio interferometric calibration, as combinations of the OS method with LS and SAGE calibration techniques. We show that the OS calibration provide a significant improvement in the execution speed compared to the conventional (non-OS) calibration algorithms. The key idea is to partition the observed data into groups of sub-observations for which the gain errors are considered to be fixed. OS-type calibrations solve for every group by iteratively updating the solutions for that group's sub-observations in an ordered sequence. The calibrations benefit from very fast computations and preserve almost the same quality as the one obtained by the non-OS calibrations. But, we must take in to account that their accuracy never becomes higher than the one of the non-OS calibration. Simulations show that OS calibration methods have considerable computational improvements compared to the conventional non-OS calibration methods.

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018

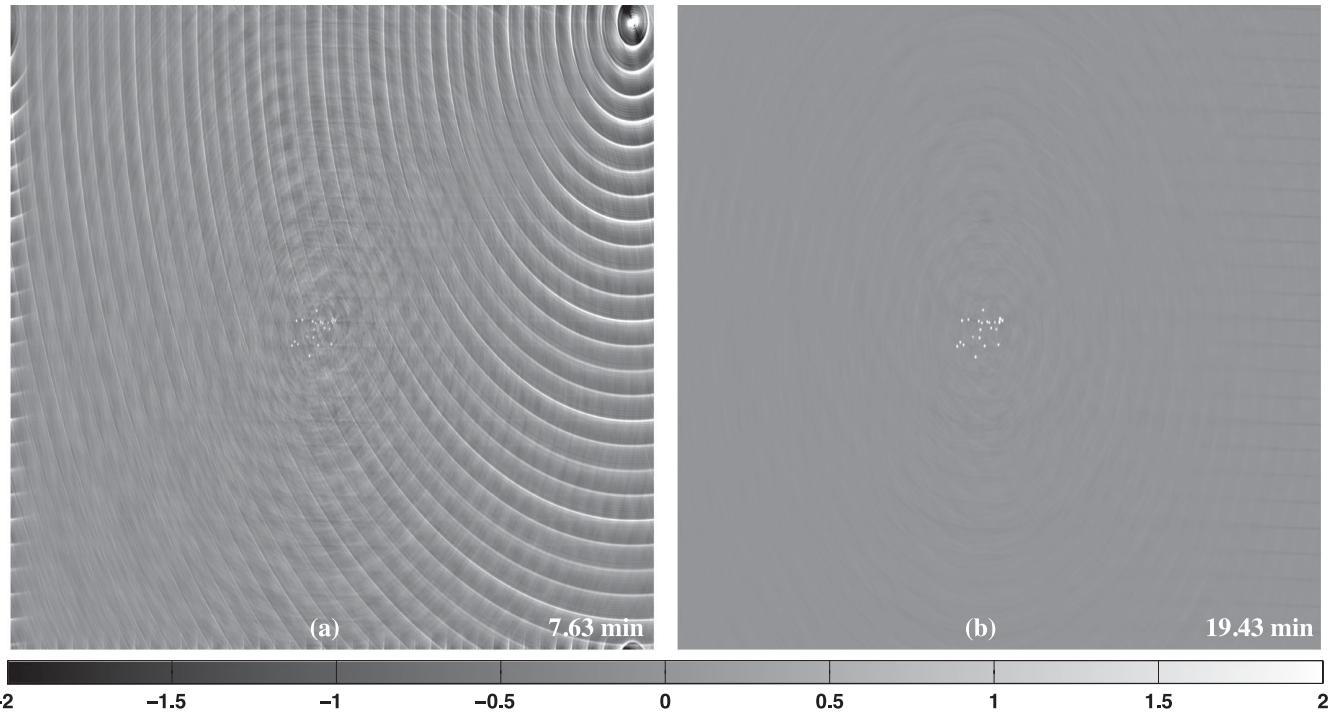

Figure 11. The residual images obtained by the LS and the OS–LS calibrations, utilizing $m = 2$ number of OS iterations and $n = 9$ number of LS iterations. The processing time is shown at the bottom right-hand corner of every image. LS calibration on averaged visibilities (a), and OS–LS (b) calibration are applied. In (a), the bright source is highly underestimated (almost not subtracted at all) and there exist severe smearing effects. This is due to the de-correlation by averaging the visibilities. However, in (b), for which OS–LS calibrations is applied on individual integrations, the bright source is perfectly subtracted and the other fainter sources are completely visible. That makes OS–LS calibration the method of choice, despite its longer execution time.

They also indicate that the OS–SAGE calibration provides a better quality results in a shorter time compared to the OS–LS calibration, as it is the case for the conventional SAGE and LS calibrations. In future work, we address a novel accuracy of calibration obtained via a hybrid of non-OS and OS calibration techniques which has a computational cost almost as cheap as the one of OS calibrations.

ACKNOWLEDGEMENTS

We would like to acknowledge the reviewer, editor, and the assistant editor for a careful review and valuable comments. S. Kazemi would also like to acknowledge NWO grant 436040.

REFERENCES

Ahn S., 2004, PhD thesis, Univ. Michigan
Born M., Wolf E., 1999, Principles of Optics. Cambridge Univ. Press, Cambridge
Burke B., Graham-Smith F., 2009, An Introduction to Radio Astronomy. Cambridge Univ. Press, Cambridge
Diestel R., 2012, Graduate Texts in Mathematics, Vol. 173, Graph Theory, 4th edn. Springer-Verlag, Heidelberg
Erdogan H., Fessler J. A., 1999, Phys. Med. Biol., 44, 2835
Feder M., Weinstein E., 1988, IEEE Trans. Acoust. Speech Signal Process., 36, 477
Fessler J. A., Hero A. O., 1993, IEEE Conference Record of Nuclear Science Symposium and Medical Imaging Conference, Vol. 3. IEEE, Piscataway, NJ, p. 1897
Fessler J. A., Hero A. O., 1994, IEEE Trans. Signal Process., 42, 2664
Hamaker J. P., 2006, A&A, 456, 395
Hamaker J. P., Bregman J. D., Sault R. J., 1996, A&AS, 117, 137
Hongqing Z., Huazhong S., Limin L., Jian Z., 2004, 11th IEEE International Workshop on Systems, Signals and Image Processing. IEEE, Piscataway, NJ
Hudson H. M., Larkin R. S., 1994, IEEE Trans. Med. Imaging, 13, 601
Kazemi S., Yatawatta S., Zaroubi S., de Bruyn A. G., Koopmans L. V. E., Noordam J., 2011, MNRAS, 414, 1656
Kazemi S., Yatawatta S., Zaroubi S., 2013, MNRAS, 430, 1457
Lampton M., 1997, Comput. Phys., 11, 110
Levenberg K., 1944, Q. Appl. Math., 2, 164
Liu L., Yin Y., Shan B., 2005, J. Image Graphics, 10, 628
Marquardt D. W., 1963, J. Soc. Industrial Appl. Math., 11, 431
Pearson T. J., Readhead A. C. S., 1984, ARA&A, 22, 97
Thompson A. R., Moran J. M., Swenson G. W., Jr, 2001, Interferometry and Synthesis in Radio Astronomy. Wiley, New York
Yatawatta S., Zaroubi S., de Bruyn G., Koopmans L., Noordam J., 2009, IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop. IEEE, Piscataway, NJ, p. 150

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article-abstract/434/4/3130/959659
by University Library user
on 05 April 2018